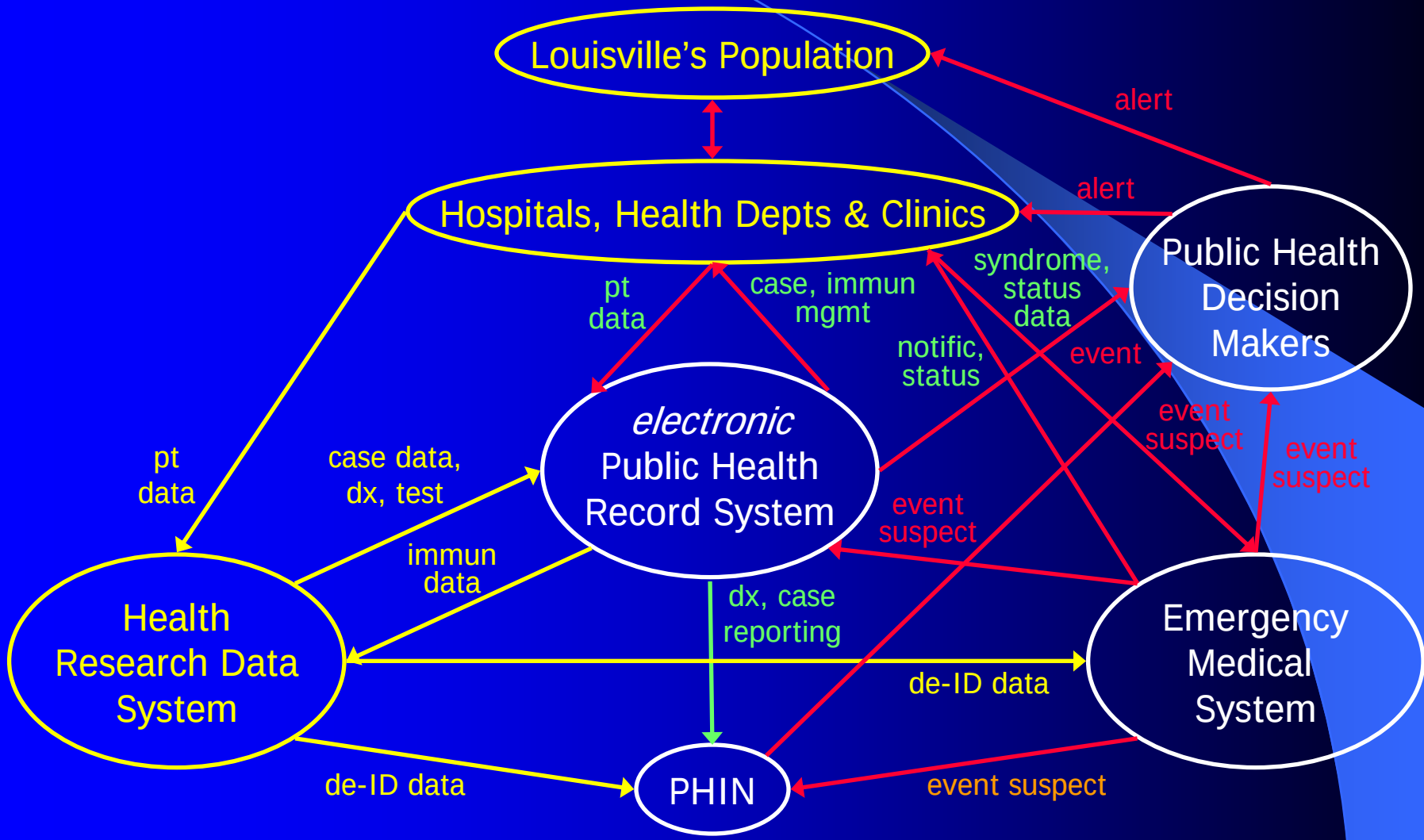
Louisville Community Surveillance Development:

Approaches to Two Problems in Population Data – ID Management and Terminology Mapping

Robert J. Esterhay, MD
Peter L. Walton, MD
School of Public Health and Information Sciences
University of Louisville

Community Surveillance Development

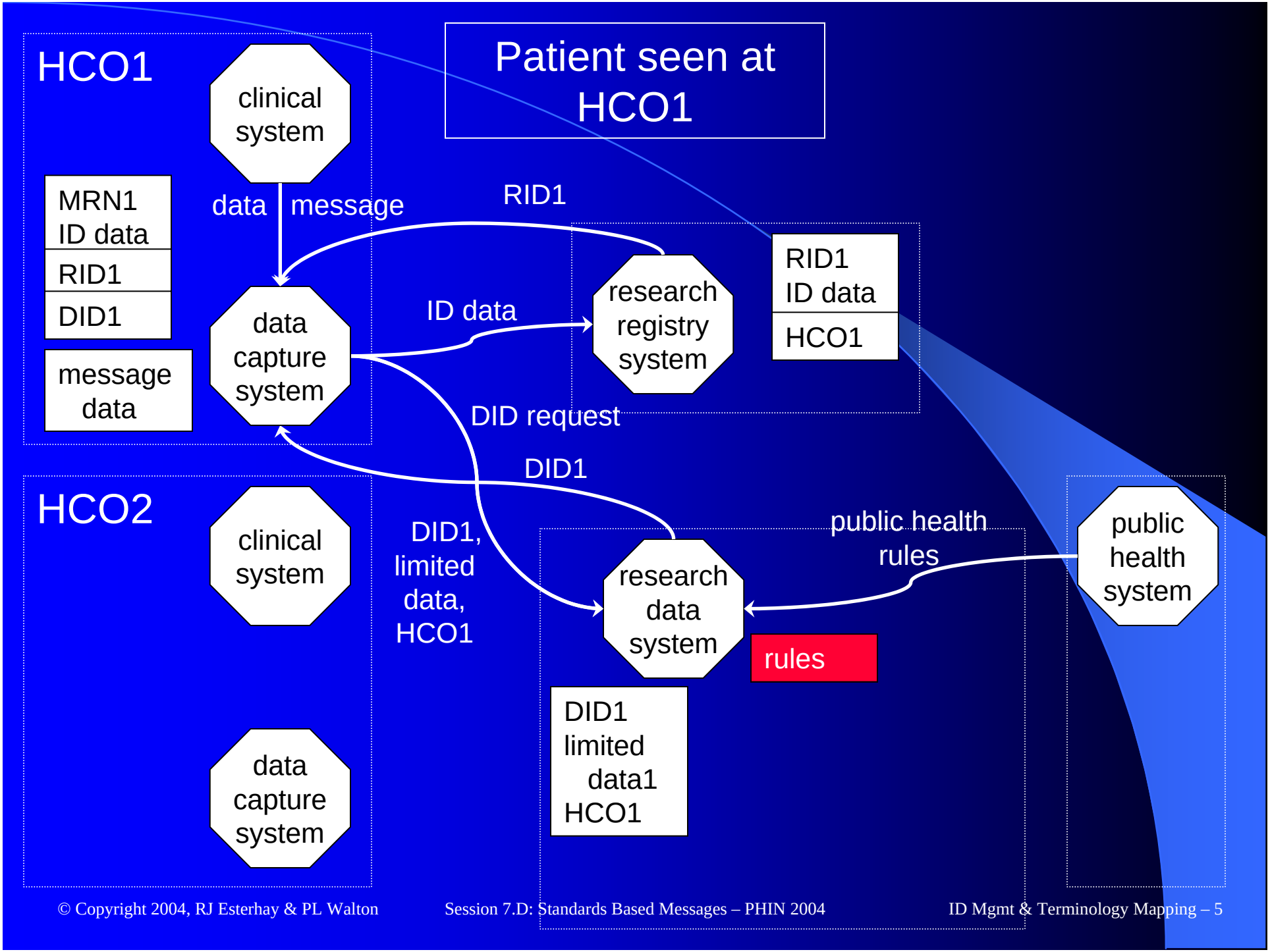


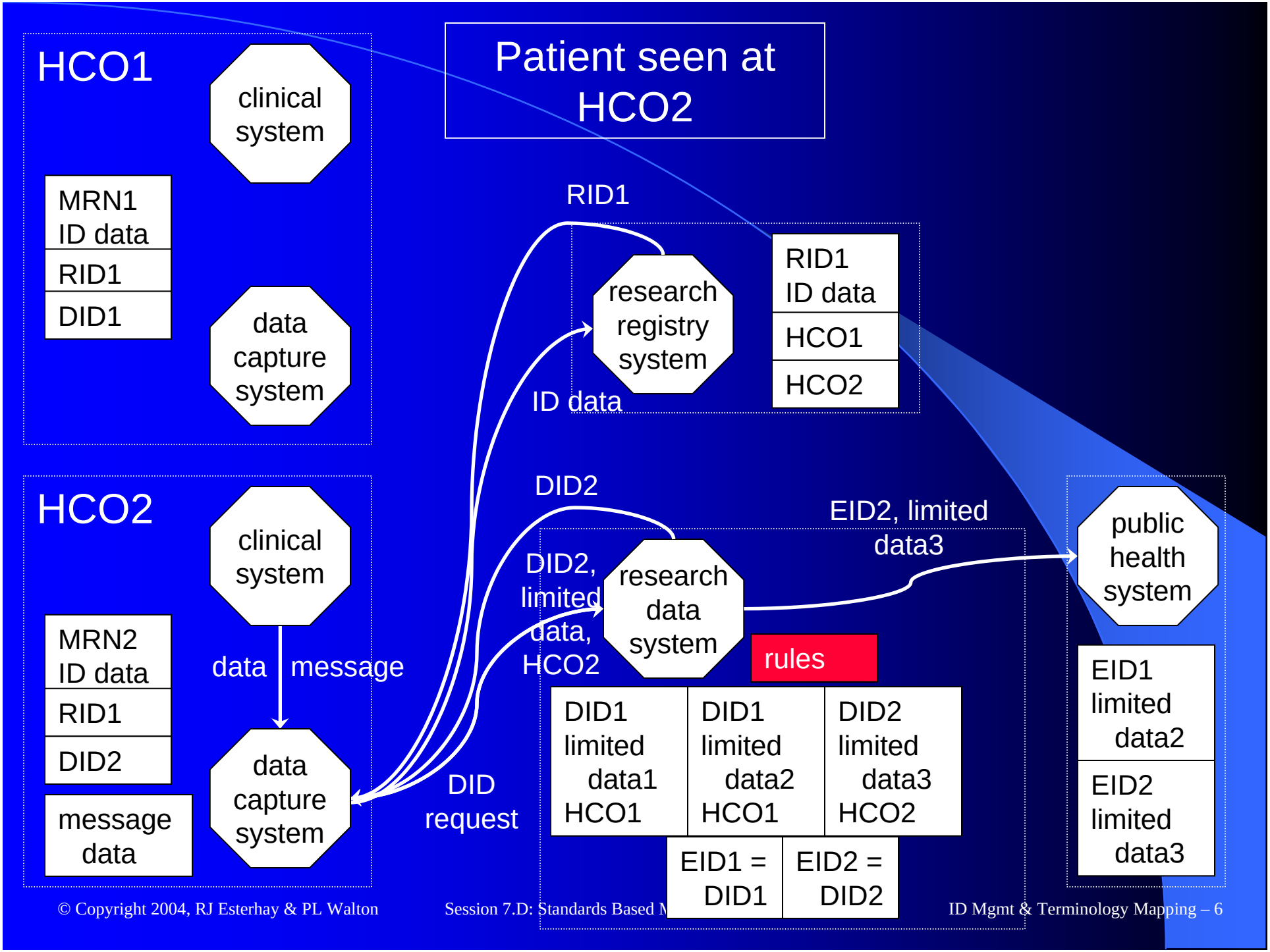
ID Management

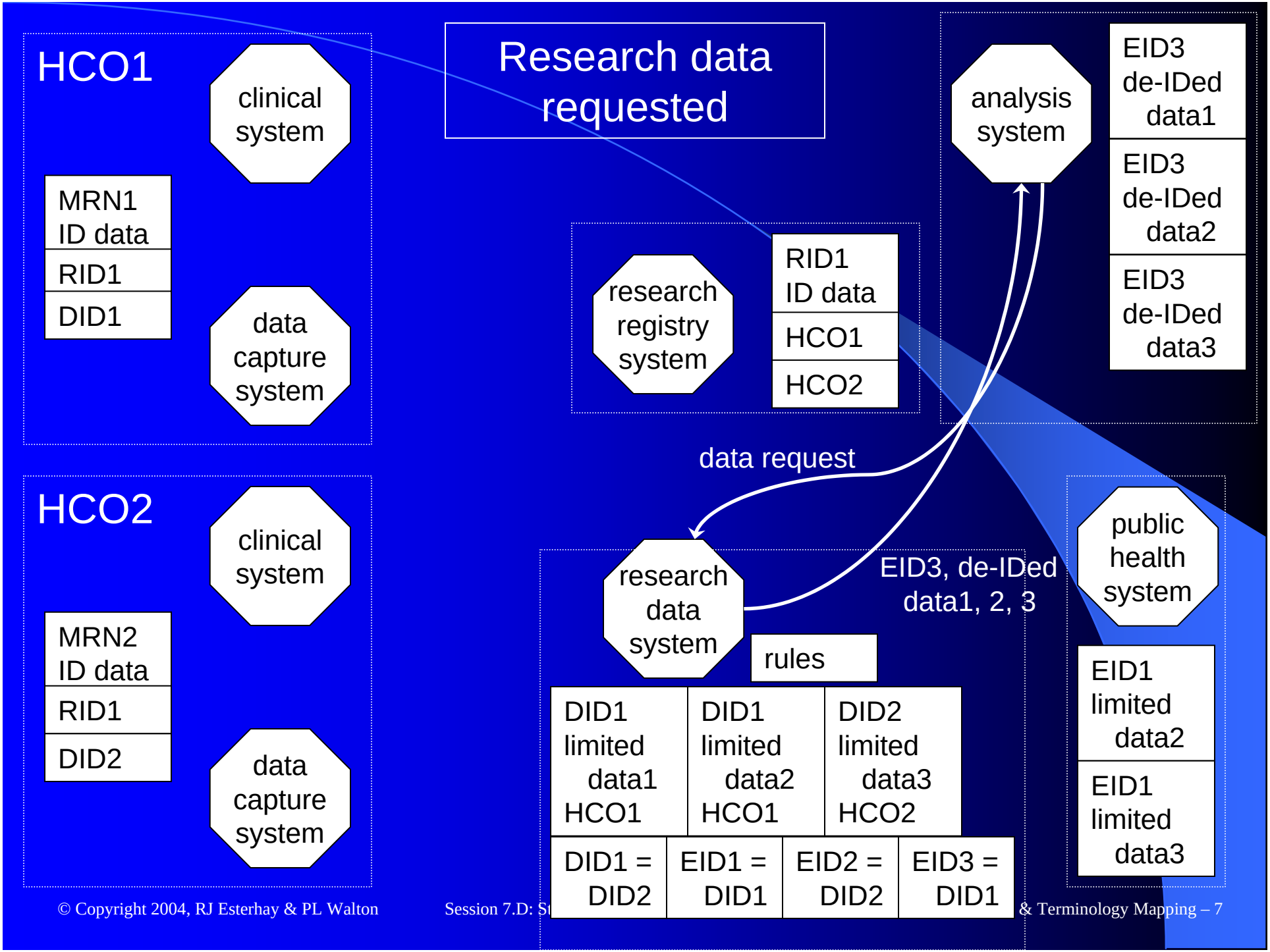
- Addresses following issues:
  - Aggregating data for an individual across multiple HCOs without violating HIPAA or requiring patient consent
  - Re-identifying de-identified data as needed
  - Labeling de-identified data for research and public health uses

Considerations

- Copy data from clinical systems into a separate system dedicated to public health and research uses.
- Use separate systems to manage IDs and to manage data.
- Have the key that links patient ID and data reside only in the institution from which data for the patient originated.


HCO1
Patient seen at HCO1
clinical system
MRN1
ID data
RID1
DID1
message data
data
message
RID1
data capture system
ID data
research registry system
RID1
ID data
HCO1
DID request
DID1
HCO2
clinical system
DID1, limited data, HCO1
public health rules
public health system
research data system
rules
DID1
limited
data1
HCO1
data capture system





Terminology Mapping

- Addresses the following issue:
  - Mapping local lab terminologies to a standard terminology with a minimum of time and effort for use in public health and research

Considerations

- If a test is mapped for one data source, it needs to be mapped (or known to be N/A) for all data sources. (Consistency rule)
- Tests need to be mapped to standard terminology only when they are needed for analysis. (“Just-in-time” rule)
- Majority of tests done are from minority of tests available. (Pareto’s “80/20” rule)

Considerations

- Mapping from a local lab test to a standard code requires knowledge of both the local lab test and the standard vocabulary.
  - In most cases neither a local lab worker nor a standard vocabulary expert has both needed sets of knowledge. (“Specific-to-general-is-harder” rule)
- Mapping from a standard lab test to a local code requires knowledge of only the local lab vocabulary.
  - Generally a local lab worker has the needed knowledge. (“General-to-specific-is-easier” rule)

“Just-in-time” Data

- Suppose a researcher or public health worker determines that the results of a lab test not in the global set are needed for analysis.
- The standard lab test, its code, name, and description are added to the global set.

New Global Mapped to Local

Global		Local Map Source1		Local Map Source2		Local Map Source3
Global(1)	=	Local1(1)	=	Local2(1)	=	Local3(1)
Global(2)	=	Local1(2)			=	Local3(2)
Global(3)	=	Local1(3)	=	Local2(3)	=	
Global(4)			=	Local2(4)	=	Local3(4)
Global(5)	=	Local1(5)	=	Local2(5)	=	Local3(5)
Global(6)	**=**	**Local1(6)**			**=**	**Local3(6)**
		unmapped		unmapped		unmapped

This mapping is done for each data source (using the "general-to-specific-is-easier" rule). Previously captured data for each source can then be reprocessed, adding "Global(6)" to each occurrence of "Local1(6)" or "Local3(6)".

Data Reprocessed

HRDS

M01:Source1
Local1(2)
= Global(2)

M02:Source1
Local1(1)
= Global(1)
Local1(5)
= Global(5)

M03:Source2
Local2(4)
= Global(4)
Local2(5)
= Global(5)

M04:Source2
Local2(8)
= null

M05:Source1
Local1(3)
= Global(3)
Local1(7)
= null

M06:Source2
Local2(1)
= Global(1)

M07:Source3
Local3(1)
= Global(1)
Local3(2)
= Global(2)
Local3(5)
= Global(5)
Local3(4)
= Global(4)

M08:Source1
Local1(2)
= Global(2)

M09:Source3
Local3(6)
= Global(6)

M10:Source2
Local2(3)
= Global(3)

After the researcher's requested test is added to the maps, the data are reprocessed. One occurrence of codes mapping to the new global code is found and mapped (shown in red).

Getting Started

- Unfortunately, in the beginning $n = 0$; that is, there are no data sources mapped into a global set and, worse, there is no global set.
- There are two ways to get started:
  - Determine the high-volume lab tests regionally or nationally (ideal).
  - Determine the high-volume lab tests from what will become data source 1 (default).
- Next the global name, code, and description for each high-volume lab test are determined. These now form the global set of lab tests.
- The global set is then submitted to one or more data sources for mapping (the “general-to-specific-is-easier” rule).

Application

- This mapping approach could be useful for public health and research, but NOT for patient care situations:
  - Patient care requires all tests to be mapped before going into production use.
  - Public health and research can tolerate limited data ("80/20" rule) and delayed availability ("just-in-time" rule).
- We are in the process of getting started (using national statistics).
- So we have not yet fully tested the approach.

Thank You